Modeling and Optimization of a Special Multistage Star Switching (SMSSS) System with SEs at Unequal Port Rates

XU Zhanqi[1], WANG Chunting[2], ZHOU Zhiqiang[3], HUANG Jiangjiang[1], MA Tao[1]

[1] State Key Lab of ISN, Xidian University, Xi'an 710071, China
[2] No.54 Institute of CETC, Shijiazhuang 050081, China
[3] Department of optical networks, FiberHome Technologies, Wuhan, 430074, China

Abstract: There has been lack of an efficient design and evaluation method for the multistage star switching (MSSS) architecture in which the ports' rates of each switching element (SE) are unequal. Thus, we identify and propose a special MSSS (SMSSS) model for the first time, where all special SEs, known as basic switching modules (BSMs), are connected hierarchically into a tree profile. Unlike the existing investigations, each BSM in this model is characterized by one high-rate port and several low-rate ports. This study focuses on the analysis, design and optimization of the SMSSS model. Moreover, we propose a novel BSM cost model which relates to its flux factor considered rarely in existing studies. Two examples are demonstrated to obtain the optimal structure parameters of the SMSSS system with a minimum overall cost. The comparison of the proposed SMSSS with similar fat tree structures indicates its relative advantages.

Keywords: multistage interconnection networks (MINs); design optimization; basic switching modules (BSMs); minimum overall cost; extended generalized fat tree (XGFT).

I. INTRODUCTION

1.1 Background

Multistage interconnection networks (MINs) have been used widely in many fields, such as traditional electronic and/or optical switching equipment, multiprocessors system, Network on Chip (NoC). The investigations on MINs can be classified into four aspects of topology, routing algorithm, performance evaluation and optimal network architecture design [1-4]. Many researches on MINs have been conducted [5-10], mainly focusing on constructing medium- and large-sized switching fabrics by interconnecting many switching elements (SEs). There has been increasing attention to the multicast [8] and/or prioritized [9] services on MINs. As a subcategory of MINs, the structure of the multi-stage star switching (MSSS) or starlike tree (also known as fat-tree (FT) in Refs. [10,11]) has the feature that the number of SEs located at each stage decreases monotonically while the speed of SEs at each stage increases as the sequence number of each stage increases.

In the application of SEs, Synchronous Digital Hierarchy (SDH) and Ethernet are the most important techniques used widely

This study focuses on the analysis, design and optimization of the special multistage star switching model. Moreover, we propose a novel basic switching module cost model which relates to its flux factor considered rarely in existing studies.

in the broadband networking of star, mesh, ring, etc. Both techniques have such an important feature that the speeds of the adjacent rate-level interfaces increase geometrically. For example, the bit rates of SDH and Ethernet are synchronous transmission module-n (STM-n, n=1,4,16,64,256) and asynchronous xG (x=0.1,1,10,100, Gigabit). Therefore, the resulting rate ratios of a level to its adjacent lower level are 4 and 10, respectively. So, it is this feature that satisfies the needs for various bandwidth granularities, thus making the MSSS structure very suitable for networking the SDH- or Ethernet-based equipment.

A special SDH or Ethernet multiplexing and switching equipment used widely in access or edge networks is usually characterized by one high-rate port and several low-rate ports. To easily describe and distinguish this special equipment from traditional SEs, we hereafter refer to the equipment of this kind as the basic switching module (BSM). As two typical examples of a BSM, one SDH Add-and-Drop Multiplexer (ADM) could have one STM-4 port and four STM-1 ports, while an Ethernet switch could have one 10GE port and 16 GE ports. Therefore, the assumption that all ports have an equal rate in Refs. [10,12] is largely different from the practical equipment employed, and degrades the values of the studies made.

Practically, such switching equipments with SDH and Ethernet ports at different rates have been used to form the networking structures stated above. Because of the needs of various bandwidth granularities in different staged (core, convergence, and access) networks and multiplexing several low-rate data services onto a high-rate one, we define the MSSS consisting of BSMs as the special multistage star switching (SMSSS), which is one of the most important networking structures and has been extensively employed in practical SDH and Ethernet networking, especially in access networks.

If the number of ports or users is assigned, we need to find the parameters that specify the SMSSS system with a minimum overall cost, i.e., three numbers of stages, BSMs at each stage and low-rate ports of a BSM. Constructing the SMSSS system with a minimum overall cost remains a bit complex since the overall cost of the system studied is closely related to some factors [4,12], such as the cost of each BSM and total number of BSMs. We propose a BSM cost model which relies not only on the number of a BSM, but also on the flux passing through it. Especially, this flux factor was rarely included.

To make the flux-based BSM cost model feasible, we propose a new model describing the SMSSS based system in which many BSMs are interconnected hierarchically into a tree-shape (detailed in Section II), and analyze the flux of each BSM. This effort makes our model closer to the practical one compared with existing studies.

Note that it is understandable for us to use a logical tree structure describing the network in which its BSMs are geographically located with the star or tree.

1.2 Related work

The FT-based topology was proposed in Ref. [13] as k-ary tree, in which the rate of an upward port of an SE at each stage is much faster than that of a downward port, and the port rate of each SE becomes greater when it approaches the root.

The extended generalized fat tree (XGFT), denoted by $XGFT(h;m_1,...,m_h;w_1,...w_h)$, was proposed in Ref. [10], where h is the height, and m_1 and w_1 are the numbers of the upward and downward ports of an SE at stage i, respectively. Many usual MINs belong to the special case of XGFT, such as [10,12,14,15] k-ary n-tree and its slimmed variants.

In the structure aspect, a topology based on the k-ary n-tree [14] has the advantages of the low latency, high bandwidth and high connectivity interconnecting end users like multiprocessors. To reduce the cost of such a network topology, Ref. [12] proposed the network based on the slimmed tree, k:k'-ary n-thin-tree, where k and k' denote the numbers of downward- and upward-port. The number

of upward-port is reduced by setting $k'<k$ instead of $k' = k$ in the k-ary n-tree. Note that the topology of both networks is regular. If we let $k' = 1$, increase the rate of an upward-port for guaranteeing the required bandwidth to avoid the congestion appearance, and use an irregular topology to construct a switching system, it is feasible that the cost of a switching system could be reduced further.

We could regard the SMSSS model as the special case of XGFT, $XGFT(h;m_1,...,m_h;1,...,1,0)$, since its BSM (except one at the highest stage) has one father node and different numbers of children nodes at different stages. Ref. [10] also addressed the topological properties of XGFT in detail. However, it did not deal with practical applications. Particularly, Refs. [4,10-12] assumed that all ports of each SE have the equal rate, which is significantly different from some application scenarios and the proposed SMSSS.

As we mentioned above, the overall cost of a switching system depends on many factors, usually including the costs of nodes, links, installation, and even maintenance [12]. It is dominated by the first factor and affected slightly by the last two factors whose contributions to the costs rely largely on application scenarios and are hard to decide. Therefore, these two factors are thus omitted usually. In a large portion of existing studies, the overall cost includes only one factor, e.g., minimizing the total number of SEs [6]. An exception [4] is that the overall cost was related to the factors of node and link, and expressed as $ax^2 + bx + c$, where x denotes the number of ports in an SE. However, determining such coefficients is complicated and the link length is also neglected. In Ref. [12], the SE's cost was assumed to be one term of ax^2, bx, or c.

Our overall cost just concerns the cost of each node or BSM, which has the forms of $bx + c$ and $ax^2 + bx + c$ for SDH and Ethernet switching (described in Section IV), respectively. We know that the number of ports of a BSM equals exactly the number of links of an SE. The link cost uncertainty in Ref. [4] is avoided in our model since the cable or fiber in China is quite cheap (just one yuan RMB per meter). Moreover, the costs of electrical cable connectors and optical transceivers are regarded to be in the cost of a BSM. Therefore, it is comparatively easy to determine the parameters of our BSM cost model.

Existing literature including Refs. [4,13] usually assumes that all ports of each SE have the same rate and thus the equal cost. This is quite impractical for the equipment with unequal port rates, e.g., for an ADM with one STM-4 port and four STM-1 ports, the cost of a high-rate port is significantly larger than that of a low-rate port. More importantly, a key factor neglected in Refs. [4,7,12] is that the cost of a switch has no relation to its throughput or port speed. For example, an 8-port Gigabit Ethernet switch may cost three times as much as one with 8-port 100M.

The criterions, which uniquely determine the optimal switching structure of the SMSSS or other fat-tree networks, usually include the number of switching modules, average number of links (or hops) sending packets from source to destination, bisection bandwidth and delay [4,10]. The defection of performance evaluations with one single criterion is obvious, for each one just represents one single character of an evaluated switching structure, e.g., the bisection bandwidth concerns mainly reliability.

There is an improved criterion in Ref. [10] whose cost of a network is characterized by the product of its diameter and degree, $d \cdot \delta$. Here, the diameter d of a network is the maximum hops among all node pairs, while the degree δ of an SE is the number of ports. Since each SE's δ is unequal in an irregular network like SMSSS, we see that it is relatively fair to use $\overline{d}$ and $\overline{\delta}$, which are the averaged d (i.e., the number of hops) and the averaged δ, respectively.

Based on the discussion above, the main contributions of this paper are as follows:

1) Through derivation from the applications scenarios that are used widely, we identify the SMSSS model for the first time, in which a BSM has one high-rate port and several

low-rate ports, i.e., the unequal port rates. Particularly, both the number of ports in each BSM and the number of BSMs may change at each stage in the SMSSS based system, which could optimally meet different requirements. However, such numbers were assumed to be fixed in previous studies as in Refs. [4,12]. To the best of our knowledge, the proposed SMSSS is so far the most accurate representation of the practical networks used widely, in particular, in access networks.

2) We propose the BSM cost model, which includes its flux or throughput as one factor, and is relatively simple and easy to decide while remaining essentially similar to Ref. [4].

3) The average diameter $\overline{d}$ and degree $\overline{\delta}$ are defined, and then $\overline{d}\cdot\overline{\delta}$, instead of $d\cdot\delta$ in Ref. [10], is used to evaluate the performance of a network, making the comparison relatively fair.

The remainder of this paper is organized as follows. Section II introduces the construction and packet switching procedure of the SMSSS model. In Section III, we make an analysis of the fluxes passing through BSMs at each stage, study the overall cost function and propose an optimization algorithm. In Section IV, we give the numerical calculation methods and optimization results via two examples. Section V compares the SMSSS with other models having similar structures and makes the multi-metric comparison.

II. Model and Switching Procedure

In this section, we first introduce the relevant switching architectures which are interconnected by BSMs and traditional SEs. We then define some parameters to describe the SMSSS model and show how to construct the SMSSS based system using BSMs in algorithm 1. We present algorithm 2 to find the root of a resulting subtree in the SMSSS. Based on algorithm 2, we discuss the packet switching by using algorithm 3.

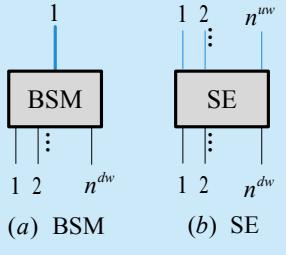

Fig.1 *Comparison* **of** *a BSM and an SE.*

2.1 BSM and networking

As shown in Figure 1, a BSM has one upward port whose rate is much larger than that of each downward port. However, the rates of the upward and downward ports of a traditional SE in Refs. [10,14] were assumed to be just the same, and the number of upward ports ($n^{uw} > 1$) used for SEs interconnection was usually equal to the number of downward ports (n^{uw}).

Suppose we are to construct a switching system with 256 ports. Figure 2 through Figure 4 present three examples in which the network structures are based on SMSSS, butterfly fat tree (BFT) [12] and Clos [16], respectively. For all three figures, we choose that the number of user ports of a BSM or an SE in the first stage equals 32 or 16. This is due to the fact that the quasi-optimal number of a single SE is 16 through 32 in Ref. [6]. Therefore, we do not consider an impractical single switch with 256 ports.

Similar to Figure 2, the rate of an upward port of each SE in Figure 3 is greater than that of a downward port. However, BFT networks like Figure 3 usually belong to regular networks. The function of each SE in Figure 3 may be the same as that of each BSM in Figure 2, while the numbers of downward ports of the BSMs at different stages could be changed to accommodate the diversified scenarios.

We note from Figure 3 that, if the number of SEs at stage 2 is reduced from 4 to 2 and the rate of each upward link is doubled, which still meets the required bandwidth of different SEs in the first stage, the BFT network could be slimmed further.

In Figure 2 and Figure 3, a packet needs to be switched only once or three times according to whether the source and destination

ports belong to the same BSM/SE at the first stage or not. However, we always need to perform switching three times in Figure 4 for any source and destination pair.

We can see that Figure 2 needs only one BSM at stage 2 while Figure 3 needs four or two SEs. Comparing these two figures, we find that the number of interconnection links is also reduced from 64 (or 32) to 4 with little loss of reliability (see more comparisons in Section V), thus reducing cost significantly in Figure 2.

2.2 Model construction

Derived from the hierarchical tree switching architecture used widely in the practical networks, the proposed SMSSS model is presented in Figure 5, in which many BSMs, represented by solid-line rectangles, are orderly connected together from the lowest stage *1* to the highest stage *N*. There are two types of nodes, leaf and non-leaf, corresponding to users and BSMs, respectively. Let $SMSSS(k,X_k,Q_k)\{k = 1,2,...,N\}$ denote the SMSSS structure, where k represents the sequence number of a stage, X_k is the number of BSMs at stage k, and Q_k denotes the number of low-rate ports of each BSM at stage k. Besides, N is the maximum number of stages.

Let $BSM(k,m)$ be each m-th BSM located at stage k. Each $BSM(k,m)$ has (Q_k+1) ports, where the Q_k low-rate ones with the same rate and the single high-rate one are on the down- and up-side of each BSM, respectively. However, a BSM at the highest stage, the N-th stage, has only Q_N low-rate ports.

Given the relevant parameters to be addressed in Section IV with two examples, $Q_k(k = 1, 2, ..., N)$ and N, it is not hard to use BSMs to construct the SMSSS model in Figure 5. Note that we do not need X_k since it can be derived from Q_k shown in Eq. (1). Alg. 1 presents the diagram of this process which is referred to as the recursive construction in Ref. [10]. We see that X_k decreases gradually to X_N=1 as $k(\leq N)$ increases, forming the multi-stage star topology.

From the leftmost side, every Q_{k+1} BSMs which are at the k-th stage and connected to the same BSM at stage k+1 constitute a cluster, including those BSMs inside the dashed-line rectangles filled with gray in Figure 5. For example, the cluster 1 consists of Q_2 BSMs starting from the leftmost side of the first stage, and all such BSMs from $BSM(1,1)$ to $BSM(1,Q_2)$ are regarded as the attachments of $BSM(2,1)$.

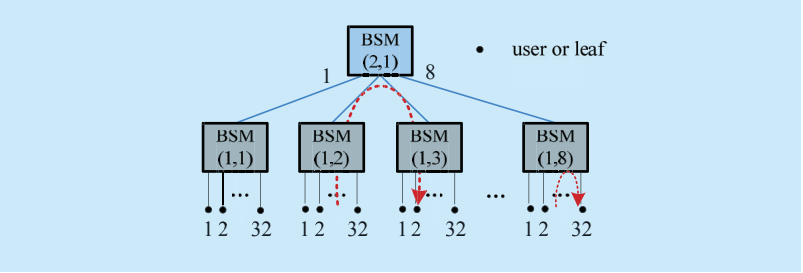

Fig.2 *SMSSS network.*

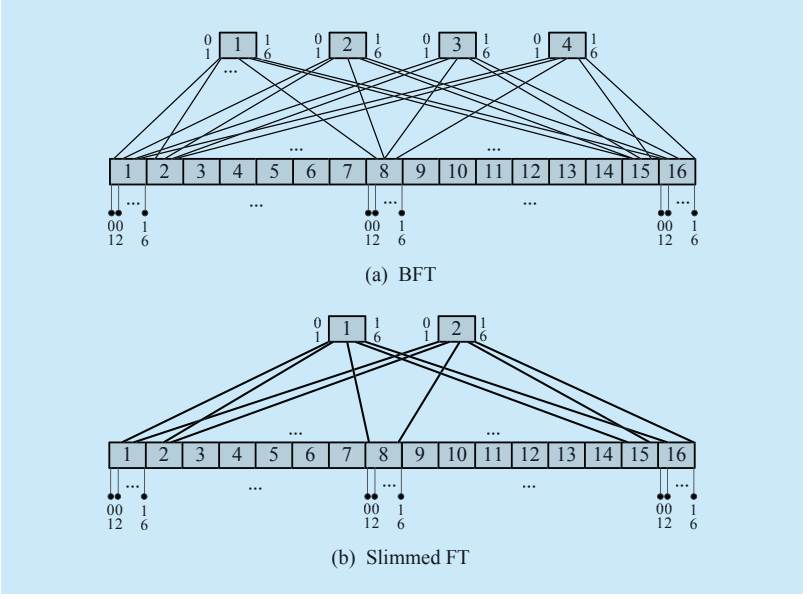

Fig.3 *BFT and slimmed FT networks.*

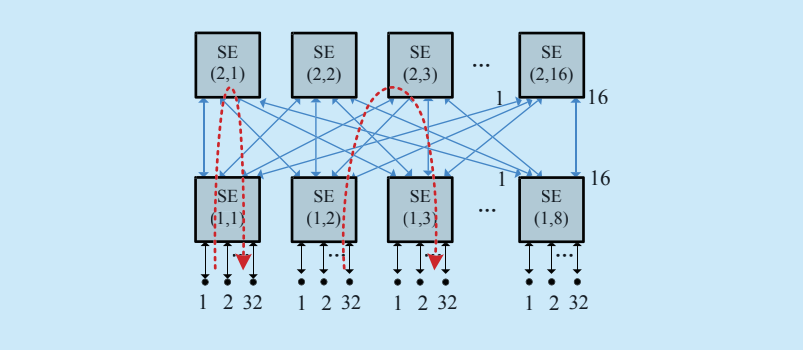

Fig.4 *Clos network (bidirectional links).*

Algorithm 1: Construct the SMSSS Model (CSM)

Steps:

1 Place one BSM with Q_N low-rate ports.

2 **for** k=N-1 to 1 **do**

3 Put $X_k(=\prod_{i=k+1}^{N} Q_i)$ BSMs in the k-th stage with each BSM having Q_k low-rate ports.

4 For all BSMs in the k-th stage and from the leftmost position, we orderly connect the high-rate port of each BSM in the k-th stage to each low-rate port of the relevant BSMs at stage k+1.

5 **end for**

2.3 Switching procedure

To prevent the switching resource from being wasted, the reasonable precondition is that, for any specific source-destination users pair, the switching should be performed at a BSM with the stage as low as possible, and thus there is only one fixed path for a packet transferred within this pair. Attention should be paid that, on the basis of the precondition mentioned and the statement in Subsection 2.1, it is impossible to use unidirectional links in Figure 5. Assuming there is an uniform traffic distribution among all users, we conclude that, as the fat tree [10] implied, the higher stage a BSM is located at, the higher the speed of its low-rate port will be.

For the easy description, let us number all low-rate ports at stage k(= 1,...,N) from the

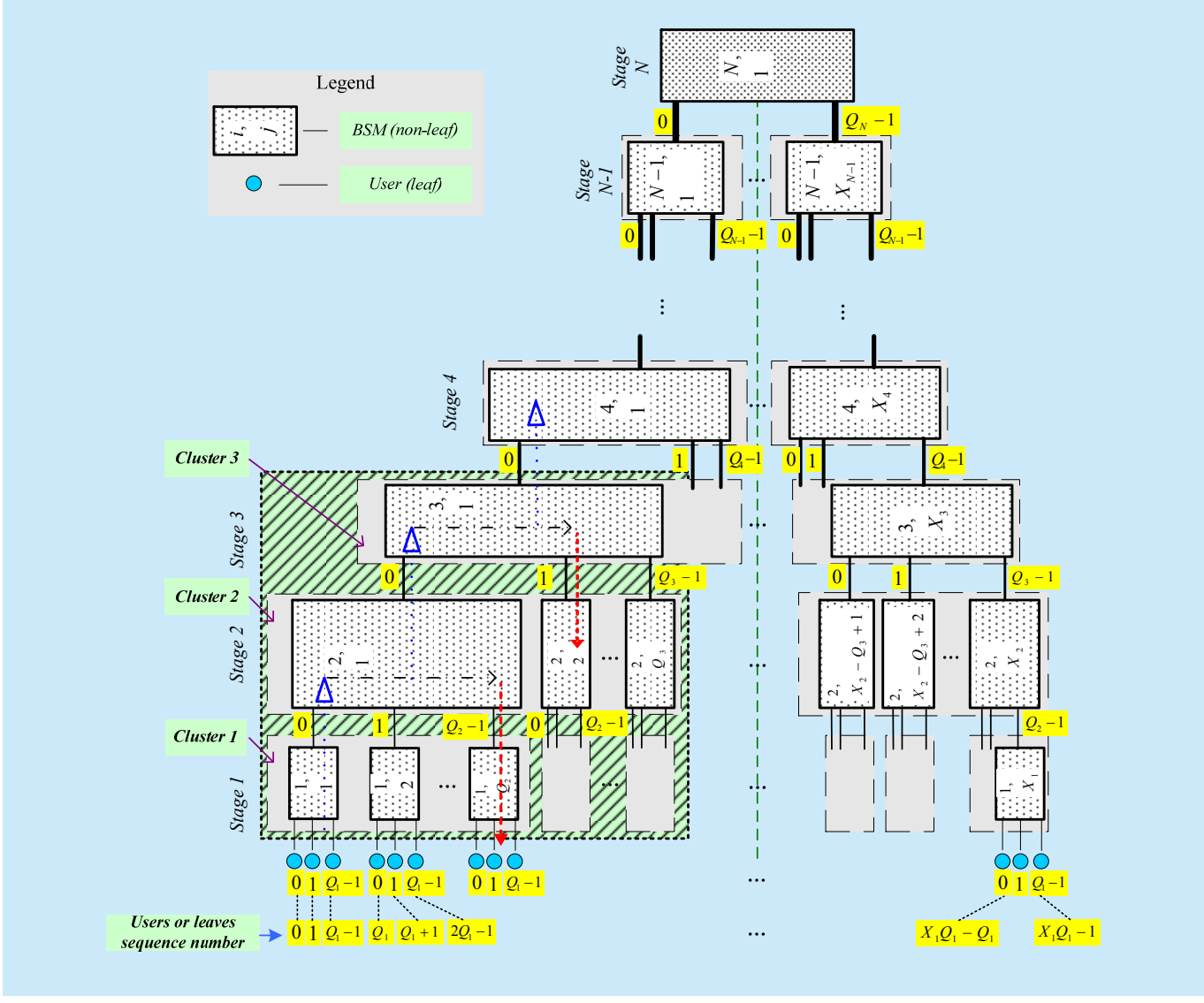

Fig.5 *Structure of the SMSSS(k,X_k,Q_k) model.*

leftmost side as (0,1,…, Q_k−1; Q_k,…, 2Q_k−1; …, $\prod_{i=k}^{N} Q_i - 1$). Note that, as shown in the downside of Figure 5, such numbers represent the sequence numbers of all leaves (users) if k=1. For any $i, j(\neq i) \in [0, \prod_{i=1}^{N} Q_i - 1]$, let $p(i, j)$ denote a constant-length packet with an input from any low-rate port of a BSM in stage 1, with the source i and destination j.

Definition 1: $Subtree_{Min}(i,j)$ is such a subtree generated from $SMSSS(k,X_k,Q_k)$ with the root of $BSM(m,n)$, which has the least possible stage sequence number m (or the lowest possible height). Besides, its attached BSMs from stage m−1 down to stage 1 exactly include the leaves i and j, i.e., $BSM(m,n)$ is the root of the resulting subtree with i and j as the leaves.

Alg. 2 presents the procedure to find the root of such a resulting subtree including the leaves i and j. We find $Subtree_{Min}(i,j)$ and $BSM(m,n)$ for any given packet $p(i, j)$, and then the model in Figure 5 works as follows. Suppose that a constant-length packet $p(i, j)$ is fed from leaf i and needs to output leaf j, and then Alg. 3 gives the procedure to perform its switching. For example, for the rectangular area filled with diagonal lines shown in Figure 5, we see that, when i and j are attached to the BSMs of this area, the switching will take place at the highest possible stage 3, 2 or 1, respectively. We can also find that the output to the high-rate port of a BSM in the upward direction is marked by the dotted-line with a triangular arrow, while the output from the low-rate port of a BSM in the downward direction is marked by the dashed-line with a small arrow. A transitional packet entering from any low-rate port of the highest staged BSM must be output to one of the remaining ports, since this BSM has no high-rate port.

We can see that, for a packet $p(i, j)$, there is a unique path in either the down to up direction or the up to down direction. The alternative to steps 5 and 6 of Alg. 3 is to run Alg. 2 for the leaves(j,i) pair, to record the path from down to up. We then use this path information in Alg. 3 in the reverse direction, i.e., from up to down.

Algorithm 2: Search the Root of a Resulting Subtree (SRRS) for leaves(i, j) pair

Input: N, $Q_k(k = 1\sim N)$, leaves(i, j) pair
Output: the smallest possible BSMs stage sequence number m
Steps:
1 $k_1 = \lfloor i/Q_1 \rfloor + 1$, $k_2 = \lfloor j/Q_1 \rfloor + 1$; Calculate the sequence numbers of BSMs to which the leaves i and j are attached, where $\lfloor x \rfloor$ equals the greatest integer not larger than x.
2 **for** m=1 to N **do**
3 **if** k_1=k_2, go to 7.
4 m=m+1.
5 $k_1 = \lfloor k_1/Q_m \rfloor + 1$, $k_2 = \lfloor k_2/Q_m \rfloor + 1$; Similar to Step 1, but at stage m.
6 **end for**
7 **return** m (transfer it to algorithm 3)

Algorithm 3: Switching of a Packet $p(i, j)$

Steps:
1 Find $Subtree_{Min}(i,j)$ by using Alg. 2 .
2 **if** m=1, perform the switching only at stage 1, go to 8.
3 **else for** k=1 to m−1 **do**
Output this packet to the high-rate port of each BSM at stage k.
4 **end for**
5 **for** k=m to 1 **do**
6 This packet in BSM at stage k is transmitted all the way down to the first staged BSM, which includes the port j of the destination.
7 **end for**
8 Output this packet to port j.

III. Model Analysis and Optimization

Since the flux or throughput of each BSM is one of the important factors affecting its cost, we study the flux of a BSM, providing the basis of optimization. We also discuss the overall switching cost of the SMSSS system, and the factors affecting the cost of a BSM. An optimization algorithm is proposed to search the three numbers describing the optimal structure of the SMSSS system designed.

3.1 Model flux analysis

3.1.1 Flux of a high-rate port

From Figure 5, we know X_N=1 and have

$$X_i = Q_{i+1}X_{i+1} \quad (i=N-1,...,2,1,0) \tag{1}$$

Note that X_0 here represents *total number of ports* (TNP) or all low-rate ports at the first stage.

Assume that the traffic input and output of users happen only at the leaf nodes, and that the input flux of every port of all BSMs at stage 1 equals one. The input flux is randomly switched to one of all (or other) ports of BSMs at the same stage in the uniform distribution. Let $f_i^{h,up}(i = 1, 2, ..., N)$ denote the upward output flux of each high-rate port in every BSM at the i-th stage, and then we have $f_1^{h,up} = (X_1Q_1 - Q_1)Q_1/(X_1Q_1 - \xi_1)$. Here, the parameter ζ_1=1 or 0 means that the destination port is exclusively or inclusively involved in the input port itself. The term $(X_1Q_1 - \xi_1)$ represents the number of all possible destination ports at stage 1, and $(X_1Q_1 - Q_1)$ is the number of all ports at stage 1 subtracting Q_1 ports of the BSM to which a packet is fed. By the analogy of $f_1^{h,up}$, we have

$$f_i^{h,up} = \prod_{j=1}^{i} (X_j - 1)Q_j^2/(X_jQ_j - \xi_j) \ (i \le N-1) \tag{2}$$

There is no high-rate port for a top BSM, and thus $f_N^{h,up} = 0$.

3.1.2 Flux of low-rate ports

Let $f_i^{l,down}(i = 1, 2, ..., N)$ denote the total downward output flux of all low-rate ports of every BSM at stage i. There is no high-rate port for a top BSM, so the total input flux from all its low-rate ports must equal the total output flux from these ports after switching. Since the number of a BSM's ports, Q_i, is just the same for each of all BSMs at stage ($i \le N-1$), and the input traffic from each port at stage 1 is uniformly assigned to all low-rate ports at the same stage, so that the total input flux from all low-rate ports of the single BSM at stage N is equally divided among all its low-rate ports, we have

$$f_N^{l,down} = f_{N-1}^{h,up}Q_N \tag{3.1}$$

As shown in BSM(1,3) and BSM(1,8) of Figure 2, there are two types of the source fluxes which generate the downward flux in each low-rate port of a BSM at stage ($i \le N-1$). In each BSM at stage (N−1) shown in Figure 5, the first portion coming from the high-rate port of a BSM is equally distributed among all its low-rate ports. Thus the first portion equals $f_{N-1}^{h,up}/Q_{N-1}$. The second portion, which is the output of the fluxes from all low-rate ports of a BSM, is just equally distributed among all low-rate ports of the BSM, $f_{N-2}^{h,up}(Q_{N-1} - \xi_{N-1})/(X_{N-1}Q_{N-1} - \xi_{N-1})$. By adding two such portions and noticing $f_{N-1}^{h,up} = f_{N-2}^{h,up}\frac{(X_{N-1} - 1)Q_{N-1}^2}{X_{N-1}Q_{N-1} - \xi_{N-1}}$ from Eq. (2), the total (downward) output flux from all low-rate ports of a BSM at stage (N−1) is

$$f_{N-1}^{l,down} = Q_{N-1}[\frac{f_{N-1}^{h,up}}{Q_{N-1}} + \frac{f_{N-2}^{h,up}(Q_{N-1} - \xi_{N-1})}{X_{N-1}Q_{N-1} - \xi_{N-1}}]$$

$$= f_{N-2}^{h,up}Q_{N-1}[\frac{X_{N-1}Q_{N-1} - Q_{N-1}}{X_{N-1}Q_{N-1} - \xi_{N-1}} + \frac{Q_{N-1} - \xi_{N-1}}{X_{N-1}Q_{N-1} - \xi_{N-1}}]$$

$$= f_{N-2}^{h,up}Q_{N-1} \tag{3.2}$$

Using Eq. (3.2) iteratively for (i = N−2, ..., 1), we could get

$$f_i^{l,down} = f_{i-1}^{h,up}Q_i \quad (i = N,...,2) \tag{3.3}$$

Eq. (3.3) shows that the total downward flux from all low-rate ports of a BSM at stage i is the upward flux from each high-rate port of a BSM at stage (i−1) multiplied by the number of low-rate ports of a BSM at stage i, i.e., for all low-rate ports of a BSM at stage i, the total downward flux equals exactly the total upward flux, satisfying the flux equilibrium.

3.1.3 Total switching flux

If $F_i(i = 1,2,...,N)$ signifies the total switching flux of all BSMs at stage i, we see

$$F_i = (f_i^{h,up} + f_i^{l,down})X_i \quad (i = 1,2...,N) \tag{4}$$

From Eq. (3.1) to Eq. (3.3) and Eq. (4), we could obtain

$$F_1 = (2 + \frac{\xi_1 - Q_1}{X_1Q_1 - \xi_1})Q_1X_1 \tag{5.1}$$

$$F_i = (2 + \frac{\xi_i - Q_i}{X_iQ_i - \xi_i})Q_iX_i\prod_{j=1}^{i-1}\frac{(X_j - 1)Q_j^2}{X_jQ_j - \xi_j}$$

$$(1 < i < N) \tag{5.2}$$

$$F_N = Q_N\prod_{j=1}^{N-1}\frac{(X_j - 1)Q_j^2}{X_jQ_j - \xi_j} \tag{5.3}$$

From Figure 5, we know

$$X_i = Q_{i+1}X_{i+1} = Q_{i+1}Q_{i+2}X_{i+2} = ... = \prod_{j=i+1}^{N} Q_j \tag{6}$$

Therefore, Eq. (5.1) to Eq. (5.3) could be expressed by using Eq. (6) as follows

$$F_1 = (2 + \frac{\xi_1 - Q_1}{\prod_{j=1}^{N} Q_j - \xi_1}) \prod_{m=1}^{N} Q_m \tag{7.1}$$

$$F_i = (2\prod_{m=i}^{N} Q_m - Q_i) \prod_{j=1}^{i-1} \frac{(X_j - 1)Q_j^2}{X_j Q_j - \xi_j}$$

$$(1 < i < N) \tag{7.2}$$

$$F_N = Q_N \prod_{j=1}^{N-1} \frac{(X_j Q_j - 1)Q_j^2}{X_j Q_j - \xi_j} \tag{7.3}$$

As stated above, we assume that the switching should be finished at the lowest possible stage, resulting in $\xi_l \equiv 1 (2 \leqslant l < N)$. If we use Eq. (6) and let ζ_1=0, Eq. (7.1) to Eq. (7.3) could be simplified by

$$F_1 = -Q_1 + 2\prod_{j=1}^{N} Q_j \tag{8.1}$$

$$F_i = Q_1 (1 - (\prod_{k=2}^{N} Q_k)^{-1})(-Q_i + 2\prod_{m=i}^{N} Q_m) \times$$

$$\prod_{j=2}^{i-1} \{(\prod_{k=j+1}^{N} Q_k - 1)Q_j^2 / (\prod_{k=j}^{N} Q_k - 1)\}$$

$$(1 < i < N) \tag{8.2}$$

$$F_N = (Q_N \prod_{j=1}^{N-1} Q_j - 1) \prod_{j=1}^{N-1} Q_j \tag{8.3}$$

3.2 Overall switching cost

As introduced in Section I, the switching cost of each BSM is closely related to several factors, such as its number of ports, port rates, switching techniques. Usually, for any BSM implementation, if either the port number Q_i or the rate of each port is increased, the implementation will become more complicated. For example, the single stage ATM switching structure fits well when the port number is not larger than a certain value (e.g., 32) [4]; however, a multistage switching structure comprehensively fits better if the port number is larger than that value.

Let $\psi(Q_i)$ be the cost of a BSM with Q_i low-rate ports at a specific speed, then the overall cost of the SMSSS system could be given by

$$C_{SMSSS} = \sum_{i=1}^{N} X_i \cdot \psi(Q_i) \cdot \gamma(i) \tag{9}$$

$$\gamma(i) = \sqrt{f_i^{l,down} + \alpha f_i^{h,up}} \tag{10}$$

where $\gamma(i)$ denotes the effect of the converted flux of a BSM at stage i on C_{SMSSS}. Parameter α is the relative importance of $f_i^{h,up}$ to $f_i^{l,down}$, and equals 0.32, 0.5, or 1. In the following examples, we let α=1.

Notice that $f_i^{l,down} + \alpha f_i^{h,up}$ exactly equals the capacity or flux of a BSM at stage i if α=1. Besides, the use of $\sqrt{}$ is due to the fact that the cost of a switch equipment or module is directly proportional to the square-root of its capacity, e.g., an 8-port 1000M Ethernet switch costs approximately three times as much as the one working at 100M. We see that Eq. (9) should be regarded as the relative overall cost because of the term $\gamma(i)$.

3.3 Optimization algorithm

The optimization computation for the SMSSS system is aiming at searching its best structure with the minimum overall cost, and deciding the detailed parameters of the SMSSS system (illustrated in Tables I through V) when the TNP in the switching system is given. Alg. 4 shows the general framework of the *Optimal Structure Searching Algorithm* (OSSA). Particularly, the condition $Q_i \geqslant Q_{i+1} (i \leqslant N-1)$ must be satisfied in step 1 since we derive Eq. (11) from Eq. (3.3) and Eq. (2)

$$\begin{aligned} f_{i+1}^{l,down} / f_i^{l,down} &= (X_i - 1)Q_{i+1} / (X_i - \xi_i / Q_i) \\ &\geq (X_i - 1)Q_{i+1} / X_i \\ &= (1 - 1/X_i)Q_{i+1} \end{aligned} \tag{11}$$

Eq. (11) indicates that the sum flux of all low-rate ports of a BSM at stage i+1 is (much) larger than that of a BSM at stage i since X_i>1 always holds if $N \geq 2$ and $Q_{i+1} \geq 4$, thereby guaranteeing the least cost of the SMSSS system.

IV. Numerical Examples

Based on the discussion in Section I, we assume $\psi(Q_i) = aQ_i^2 + bQ_i + c$ and get it by col-

Algorithm 4: Optimal Structure Searching Algorithm (OSSA)

Input: *TNP*, ζ_1 and possible Q_i numbers allowed.

Output: the optimal structure S_{opt} with the minimum overall cost.

Steps:

1 Decompose the *TNP* with the combination of different Q_i numbers allowed and get all feasible structures.
2 **for** each feasible structure **do**
3 Calculate its flux at each stage by Eq. (2), Eq. (3), and Eq. (7).
4 Calculate its overall cost by Eq. (9).
5 **end for**
6 Compare C_{SMSSS} of all feasible structures, and get S_{opt} with the minimum C_{SMSSS} .
7 Record $f_i^{h,up}(i \le N-1)$ by Eq. (2) for each BSM.

lecting the price lists of the relevant products and using the data-fitting method.

4.1 SDH case

In the first case, a BSM in the SMSSS is replaced by an SDH ADM. By using data-fitting and inquiring the FiberHome Technology Ltd., one of the main SDH manufacturers in China, $\psi(Q_i)$ could be represented in Chinese Yuan by

$$\psi(Q_i) = 6770Q_i + 115090 \quad (12)$$

where $4 \leqslant Q_i \leqslant 16$, and we should state that Eq. (12) already includes the cost of one high-rate port inside this SDH ADM equipment.

Table I gives the optimization results of the SMSSS system at different TNPs, $2^i (i = 5, 6, ..., 9)$. Each number of the feasible structures is obtained by decomposing the port number of the SMSSS system with the probable numbers of ports for each BSM, which are equal to $2^j (j = 2, 3, 4)$ and typically used in practice.

We see that the *number of optimal switching stages* (NOSS) changes from two to three when *TNP* increases from 32 to 512. Each numbers pair in *optimal switching structures* (OSS) represents the numbers of BSMs and ports for each BSM at stage *i*, e.g., for 128 ports, {8,(1/16)} means 8 BSMs with each having 16 downward ports and one upward port at first stage, and {1,(0/8)} symbolizes one BSM with 8 downward ports at second stage. From Eq. (12), we find that its constant term is much larger than the coefficient of its linear term. Thus, to lower $\psi(Q_i)/Q_i$, the Q_i trends reasonably to be larger or to be the upper bound, 16.

4.2 Ethernet case

The second example for the proposed SMSSS is to construct a multistage Ethernet switching system in which an Ethernet switch works as a BSM. Similar to the SDH cases, we use data-fitting again and get Eq. (13) by collecting the commercial prices of the Ethernet switch products from three Chinese companies, TP-LINK, H3C and Digital China

$$\psi(Q_i) = 0.088q_i^2 + 12.24q_i + 25.44 \quad (13)$$

We define

$$q_i \triangleq Q_i + \beta . f_i^{h,up} \big/ (f_i^{l,down} / Q_i) \quad (14)$$

where the term $f_i^{h,up}/(f_i^{l,down}/Q_i)$ represents the ratio of the upward flux of the single high-rate port to that of each low-rate port of a BSM at stage *i*.

Unlike the former SDH case in Eq. (12), we find that the prices from such three companies for the same switch differ greatly, and thus we use the averaged price over three companies. The majority of the prices for Ethernet switches, which only have the low-rate ports with the same number of ports and equal port-rate, are used to derive Eq. (13) at β=0 via data-fitting. By collecting the prices of the remaining Ethernet switches with one high-rate port and a variable number of low-rate ports, we could find that the cost of one high-rate port of an Ethernet switch is approximately equivalent to that of $\beta \cdot f_i^{h,up}/(f_i^{l,down}/Q_i)$ low-rate ports. The parameter β is the conversion ratio, and equals $\sqrt{10}/10 \approx 0.32$ in the following examples, since we know the cost of one GE port equals $\sqrt{10}$ if we assume that the cost of one 100M port is 1. Eq. (13) complies with the quadratic polynomial of $ax^2 + bx + c$ in Ref. [4] as stated in Section I.

Since a low-rate port of each BSM at stage $i(\leqslant N-1)$ is connected to the relevant high-rate port of a BSM at stage (i−1), we have Eq. (15) by using Eq. (2)

$$q_i = Q_i + \beta \cdot (X_i - 1) Q_i^2 / (X_i Q_i - \xi_i) \quad (15)$$

Applying Eq. (13) and Eq. (15) to Eq. (9), Table II gives the optimized results of the SMSSS system with the Ethernet switches and different TNPs, $2^i (i = 5, 6, ..., 11)$. The allowed numbers of ports for each BSM equal $2^j (j = 2, 3, ..., 6)$, which are usually employed in practical Ethernet switches.

Different from Eq. (12), Eq. (13) shows that it has a quadratic term and that its constant/(coefficient of the linear term) ratio is much less than that of Eq. (12), so that a smaller Q_i performs better. This is validated in Table II if we compare Table II with Table I for the same TNPs except for the 32-port case. However, if we increase β to 1, the NOSSs in Table II will change only in the 32-port case, from two to one. This is because from Eq. (15), a larger β results in a larger q_i and thus a higher cost.

To examine the impact the speed of each low-rate port of a BSM on the OSS and NOSS, we change Eq. (10) to

$$\gamma(i) = \sqrt{F_i / (X_i q_i)} \tag{16}$$

The relevant results are shown in Table III. Comparing to Table II, we see that the NOSS in Table III decreases by two for 1024 ports while it decreases by one for the rest of all six different TNPs. We may conclude that the larger q_i in Eq. (15) will result in the lower cost. However, Eq. (13) shows that $\psi(Q_i)$ increases when q_i becomes larger. Therefore, the adequate q_i could make Eq. (9) minimum

Table I *Optimization results under different TNPs (SDH)*

		Optimal switching structures				
Total number of ports (TNP)		32	64	128	256	512
Number of feasible structures		1	3	2	4	3
Number of optimal switching stages (NOSS)		2	2	2	2	3
Number of BSMs		5	5	9	17	37
Optimal switching structures (OSS) at stage *i*	*i*=3					1,(0/4)
	i=2	1,(0/4)	1,(0/4)	1,(0/8)	1,(0/16)	4,(1/8)
	i=1	4,(1/8)	4,(1/16)	8,(1/16)	16,(1/16)	32,(1/16)

Table II *Optimization results under different TNPs (Ethernet, Eq. (10))*

		optimal switching structures						
Total number of ports (TNP)		32	64	128	256	512	1024	2048
Number of feasible structures		2	4	3	6	6	9	9
Number of optimal switching stages (NOSS)		2	2	3	3	3	4	4
Number of BSMs		5	9	21	37	73	149	293
Optimal switching structures at stage *i*	*i*=4						1,(0/4)	1,(0/4)
	i=3			1,(0/4)	1,(0/4)	1,(0/8)	4,(1/4)	4,(1/8)
	i=2	1,(0/4)	1,(/08)	4,(1/4)	4,(1/8)	8,(1/8)	16,(1/8)	32,(1/8)
	i=1	4,(1/8)	8,(1/8)	16,(1/8)	32,(1/8)	64,(1/8)	128,(1/8)	256,(1/8)

Table III *Optimization results under different TNPs (Ethernet, Eq. (16))*

		Optimal switching structures						
Total number of ports (TNP)		32	64	128	256	512	1024	2048
Number of feasible structures		2	4	3	6	6	9	9
Number of optimal switching stages (NOSS)		1	1	2	2	2	2	3
Number of BSMs		1	1	5	17	17	33	69
Optimal switching structures at stage *i*	*i*=3							1,(0/4)
	i=2			1,(0/4)	1,(0/16)	1,(0/16)	1,(0/32)	4,(1/16)
	i=1	1,(0/32)	1,(0/64)	4,(1/32)	16,(1/16)	16,(1/32)	32,(1/32)	64,(1/32)

when we are applying Eq. (13) and Eq. (16) to Eq. (9). Since the 16- and 32-port BSMs are used in most scenarios in Table III, the result suggests that, for different factors affecting the cost of the Ethernet switch, the speed of each low-rate port of a BSM in Eq. (16) would be more reasonable compared to the capacity of a BSM in Eq. (10). The results in Table III also accords with Ref. [4] that 16- and 32-port SEs or BSMs are nearly optimal.

V. COMPARISION

To show the validity of the proposed SMSSS, we compare the cost of the SMSSS with that of the XGFT-based system whose SEs are with equal port rates (EPR). We denote such a system as $XGFT_{EPR}$. This comparison is relatively rational because XGFTs cover the popular trees, such as k-ary n-tree, butterfly FTs, and slimmed FTs (i.e., k:k'-ary n-thin-tree [12]). Since the SDH equipment whose ports rates are equal has been found rarely in applications, we only discuss the Ethernet case. In the second part, the multi-metric comparison is made between the SMSSS and closer $XGFT_{EPR}$.

5.1 Comparison with *XGFT*s at equal port rates

If a system is constructed on the basis of $XGFT_{EPR}$, Eqs. (9) through (15) must be changed to accommodate this EPR scenario, making the comparison fair

$$C_{XGFT_{EPR}} = \sum_{i=1}^{N} X_i \cdot \psi(Q_i) \cdot \gamma_{BFT}(i) \tag{17}$$

$$\gamma_{XGFT_{EPR}}(i) = \sqrt{n_i^{dw} + n_i^{uw}} \tag{18}$$

where $\gamma_{XGFT_{EPR}}(i)$ is similar to $\gamma(i)$ in Eq. (10), n_i^{dw} denotes the number of downward ports of an SE at stage i, and n_i^{uw} represents the number of upward ports which is to be used in the same SE at stage i, and equals $\lceil f_i^{h,up} \rceil$.

Eq. (13) is also valid, but q_i must be changed to

$$q_i \triangleq n_i^{dw} + m_i^{uw} \tag{19}$$

where m_i^{uw} is the number of upward ports of an SE at stage i, and satisfies $m_i^{uw} \leqslant n_i^{dw}$ since the total flue of all upward ports of an SE will never be larger than that of all its downward ports [12]. Applying $X_j Q_j >> 1 \geqslant \xi_j$ to Eq. (2), we get

$$f_i^{h,up} = \prod_{j=1}^{i} (1 - 1/X_j) Q_j \approx \prod_{j=1}^{i} Q_j \tag{20}$$

Therefore, it is reasonable that we let $m_i^{uw} \approx n_i^{dw}$ with the constraint of $q_i = 2^j$ (j=2,3,..., 6). In such a case we may leave a few upward ports unused. However, those unused ports must be included in the cost of an SE since such unused ports could not be removed from an Ethernet switch. Note that the number of unused ports, $m_i^{uw} - \lceil f_i^{h,up} \rceil$, is not included in Eq. (18), so we equivalently lower the cost of an SE. In addition, corresponding to Eq. (16), we select $\gamma(i)$ based on the assumption that the flux fed to each port equals one

$$\gamma(i) = 1 \tag{21}$$

Analogous to Tables II and III, Tables IV and V give the information related to the optimal switching structures based on the $XGFT_{EPR}$ when using Eq. (18) and Eq. (21), respectively. Note that the allowed n_i^{dw} just equals the range of Q_i, $2^i (i = 2, 3, 4, 5, 6)$, so the allowed number of downward ports for a BSM has the same range as that for an SE. We can see that the NOSS in Table IV is decreased by one compared to that in Table II, and the corresponding NOSS reduction in Table V is only with TNP=2048 when compared to Table III.

Shown in Figure 6 is the total cost comparison of optimal switching structures listed in Tables II through V with the logarithmic vertical-axis. The results marked a) and c) are based on Eq. (10) at (α = 1, β= 0.32) and Eq. (16), while results b) and d) correspond to those by using Eq. (18) and Eq. (21). Compared to the case b), the cost for the SMSSS in a) is decreased by 4.7% if TNP=32, and lowered by 36.3% to 63.6% when TNP increases from 64 to 2048. For the cases c) and d), such cost reduction ratios will be about 40% to 43% except TNP=32 and 64.

Such large cost reduction ratios result from the fact that, as shown in Tables II through

IV, the number of BSMs is greatly reduced when compared to the number of SEs in the $XGFT_{EPR}$. For example, if we assume that the rate of each downward port in first stage is 10M and TNP=2048, Table III shows that the SMSSS will need 64 switches with 32-port at 10M, 4 switches with 16-port at 100M, and one switch with 4-port at 1000M. However, Table V indicates that the $XGFT_{EPR}$ will need 96 switches of 64-port at 10M.

5.2 Multi-metric comparison

Because of the specialty of the BSMs used in the SMSSS, we try to find some FTs which are very closer to the SMSSS and evaluate the performance.

With the criterions and parameters defined in Subsection 1.2, we only try to compare the SMSSS with other models having similar structures. It is because, as stated in Section I, a BSM is distinguished from other models by its unequal port-rate and a variable number of BSMs at different stages.

Figure 7 and Figure 8 present the performance comparison of the SMSSS shown in Table III with the $XGFT_{EPR}$ by five criterions at TNP=256 and 1024, respectively. Here, $XGFT_{EPR}(N, n_i^{dw}, n_i^{uw})$ denotes a fat-tree in which n_i^{dw} and n_i^{uw} are stated in Figure 1. The

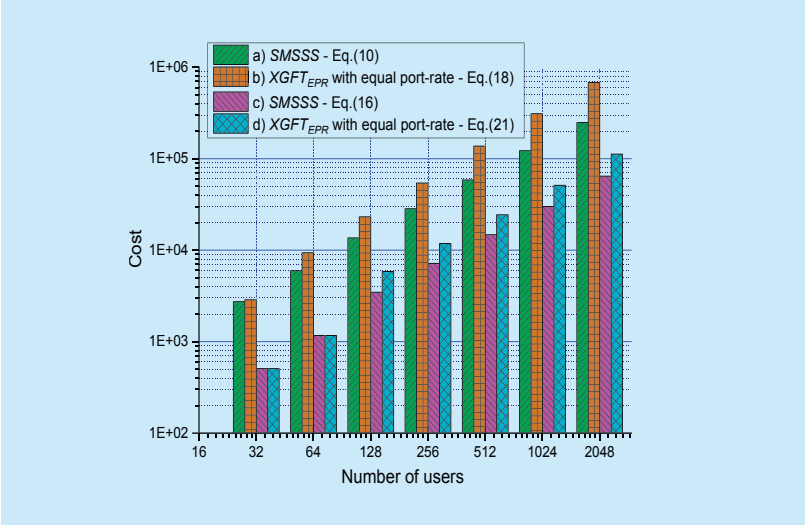

Fig.6 *Total cost comparison of the SMSSS and* $XGFT_{EPR}$

selection of n_i^{dw} and n_i^{uw} for two FTs must guarantee that the summed capacity of all upward ports of an SE must not be less than the summed flue, which is input from all its downward ports and needs to be switched to the remaining SEs. For a case that the service is distributed uniformly among all users, suppose $\xi_1 = \xi_2 = 1$ with X_0 defined in Subsection 3.1, then we have

$$n_1^{uw} \geqslant (1 - (n_1^{dw} - 1)/(X_0 - 1)) \cdot n_1^{dw} \qquad (22.1)$$

$$n_2^{dw} \geq n_1^{uw} \qquad (22.2)$$

Table IV *Optimization results for* $XGFT_{EPR}$ *under different TNPs (Ethernet, Eq. (18))*

		Optimal switching structures						
Total number of ports (TNP)		32	64	128	256	512	1024	2048
Number of feasible structures		3	5	7	12	18	29	45
Number of optimal switching stages (NOSS)		1	1	2	2	2	3	3
Number of SEs		1	1	36	40	72	416	576
Optimal switching structures at stage *i*	*i*=3						32,(0/32)	64,(0/32)
	i=2			4,(0/32)	8,(0/32)	8,(0/64)	256,(4/4)	256,(8/8)
	i=1	1,(0/32)	1,(0/64)	32,(4/4)	32,(8/8)	64,(8/8)	128,(8/8)	256,(8/8)

Table V *Optimization results for* $XGFT_{EPR}$ *under different TNPs (Ethernet, Eq. (21))*

		Optimal switching structures						
Total number of ports (TNP)		32	64	128	256	512	1024	2048
Number of feasible structures		3	5	7	12	18	29	45
Number of optimal switching stages (NOSS)		1	1	2	2	2	2	2
Number of SEs		1	1	24	40	48	80	96
Optimal switching structures at stage *i*	*i*=2			8,(0/16)	8,(0/32)	16,(0/32)	16,(0/64)	32,(0/64)
	i=1	1,(0/32)	1,(0/64)	16,(8/8)	32,(8/8)	32,(16/16)	64,(16/16)	64,(32/32)

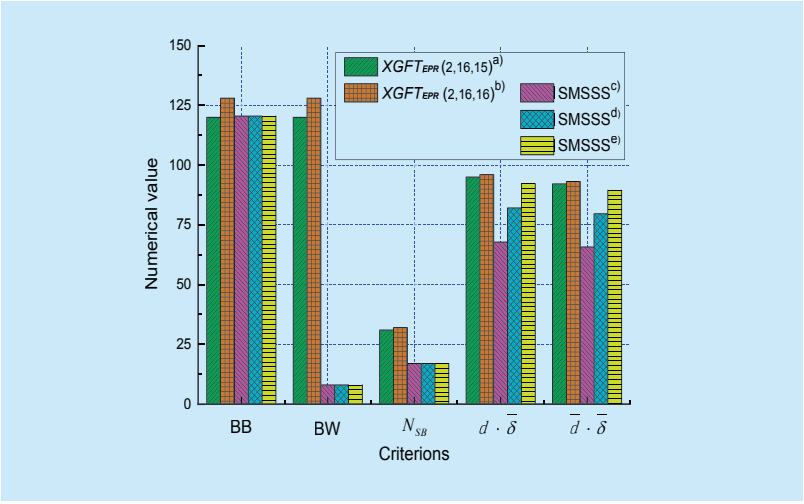

Fig.7 *Comparison of five criterions (256 user-port).*

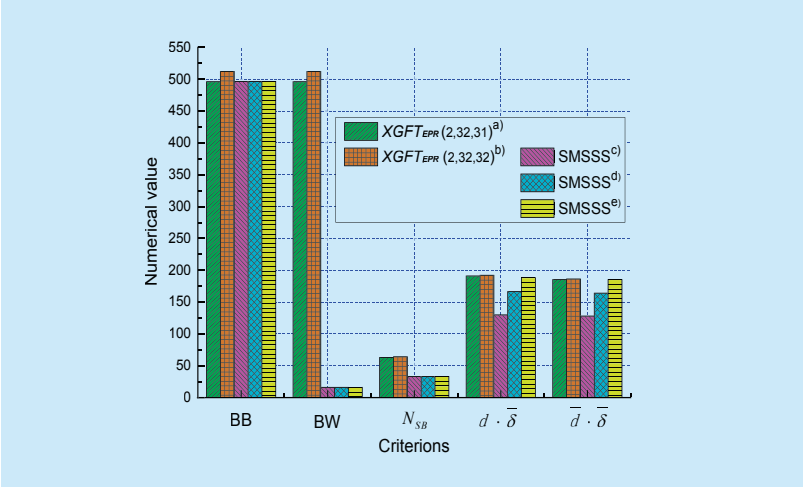

Fig.8 *Comparison of five criterions (1024 user-port).*

Attention should be paid to the fact that the minimum n_i^{itw} in the scenarios marked by the superscript a) in both figures should be 15.06 and 31.03 when strictly satisfying Eq. (22.1). Therefore, we still list them just for evaluation.

In Figure 7 and Figure 8, the switching structures of each three scenarios (two $XGFT_{EPR}$-based FTs and one SMSSS) have two stages, and each SE in two fat-tree cases has 16 or 32 downward ports, respectively. Note that the definitions of d,$\overline{d}$ and $\overline{\delta}$ can be found in Subsection 1.2, and each user link at the lowest stage (i.e., connecting a user to SEs) is included when calculating d and $\overline{d}$. Contrarily, each user endpoint is not included when calculating $\overline{\delta}$. The remaining notations for the performance parameters are as follows.

BB: the bisection bandwidth [8] which is a bandwidth across the smallest cut that divides a network into equal halves.

BW: the bisection width which refers to the number of links, not the bandwidth.

N_{SB}: the number of all SEs in FTs or BSMs in SMSSS.

For the results signed by c), we assume that the degree of a high-rate port equals one, or equivalently, q_i defined in Eq. (15) is equal to Q_i+1. To make the comparison relatively fair, the calculation of $\overline{\delta}$ for the results marked by d) and e) is based on Eq. (15) at β=0.32 and 0.5, respectively. Equivalently, we assume that the degree of a high-rate port of a BSM at stage i is

$$\omega(i)=\begin{cases}1 & case\ c)\\ 0.32\cdot(X_i-1)Q_i^2/(X_iQ_i-\xi_i) & case\ d)\\ 0.5\cdot(X_i-1)Q_i^2/(X_iQ_i-\xi_i) & case\ e)\end{cases} \tag{23}$$

Note that both figures do not include $\overline{d}$ since it is just the same for each of three scenarios, and equals 3.88 and 3.94 when TNP=256 and 1024, respectively. In both figures, the legends of $XGFT_{EPR}$-based case b) and SMSSS case c) are intentionally placed on horizontal positions since the numbers of downward ports of each BSM/SE in both stages for two structures are identical.

We find from both figures that, compared with $XGFT_{EPR}$-based FTs, the SMSSS has the smallest *BW* and N_{SB}. We can also see that, compared to the $XGFT_{EPR}$ case a) in two figures, the *BB* value of each SMSSS case is a bit larger, increasing by 0.4% and 0.1%, respectively. However, compared to the $XGFT_{EPR}$ marked b) in both figures, the corresponding *BB* value of each SMSSS case decreases by 5.9% and 3.0%, respectively. The BW ratios of the SMSSS to $XGFT_{EPR}$(2,16,16) and $XGFT_{EPR}$(2,32,32) are just 6.7% and 3.4%, which show that the SMSSS has a lower reliability in terms of the number of bisected links when compared to the two $XGFT_{EPR}$-based FTs. However, the high reliability of the

current equipment can alleviate this defection. In addition, as shown in Figure 6, the SMSSS can largely reduce the cost of a switching system due to the large reduction of the number of BSMs used.

For the comprehensive and more reasonable criterions of both $d\cdot\overline{\delta}$ and $\overline{d}\cdot\overline{\delta}$, SMSSS remains the least. We do not consider the case c) which outperforms $XGFT_{EPR}$-based FTs greatly and may be relatively unfair, since the degree of a high-rate port is assumed to be equal to one. Compared to the $XGFT_{EPR}$(2,16,16) and $XGFT_{EPR}$(2,32,32) , each value of $\overline{d}\cdot\overline{\delta}$ in cases d) and e) of Figure 7 decreases by 14.5% and 3.9%, respectively. In Figure 8, the corresponding $\overline{d}\cdot\overline{\delta}$ for each of these two cases reduces by 11.9% and 0.5%.

The performance metrics above demonstrate that the SMSSS has an obvious advantage. It is the structure character of a BSM in the SMSSS model, one high-rate port and several low-rate ports, that makes all conclusions above rational. On the other hand, the very small *BW* of the SMSSS shows its low tolerance in theory. However, the high reliability of equipment, which has been used widely in practical networks and is the implementation example of our SMSSS model, outweighs this shortcoming.

VI. Conclusions

In this paper we have presented the SMSSS model whose switching structure is quite different from the usual MINs and treelike structures [1-15]. The proposed SMSSS features that its BSM has one high-rate port and several low-rate ports, and that many BSMs are interconnected to a specific multistage star or tree-like structure. Therefore, the proposed SMSSS, which could be regarded as the most slimmed fat tree (MSFT) with SEs at unequal port rates, is suitable for the optimized design of the networking equipment as the structure stated above, in particular, the broadband SDH and Ethernet switching equipment that have been used widely in access networks. It is also possible to use SMSSS model in the other areas, such as the switch fabric with large ports, data centers and NoC.

Unlike the studies in the existing literature [10,12], we consider the switching flux of every stage as an important factor affecting the cost of a BSM in the proposed SMSSS model. Given the TNP of a switching system and the cost factors of a BSM, it is feasible to find a switching scheme with a minimum overall cost. We find that the flux passing through each port is more rational than the flux passing through the SE or BSM since the NOSSs in both SMSSS and FTs could be decreased by at the least one, increasing the reliability relatively.

Similar to the traditional performance criterions, Eq. (9) could also represent some scenarios studied already. We can assume that the overall cost just relies on the number of BSMs if $\psi(Q_i) = \gamma(i) = 1$, and the numbers of BSMs and ports of each BSM when $\gamma(i) = 1$.

One of our next works is to evaluate the SMSSS performance via simulation at different service patterns [17].

Acknowledgement

This study is supported in part by the National High-Tech Development Project (2012AA01A505), Key Issues of Terabit PTN Equipment R&D from the Ministry of Industry and Information Technology.

References

[1] GAROFALAKIS J, STERGIOU E. An Analytical Model for the Performance Evaluation of Multistage Interconnection Networks with Two Class Priorities[J], Future Generation Computer Systems, 2013, 29(1):114-129.

[2] YUAN Xin, MAHAPATRA S, LANG M, at al. Static Load-Balanced Routing for Slimmed Fat-Trees[J], Journal of Parallel and Distributed Computing, 2014, 74(5): 2423–2432.

[3] GAROFALAKIS J and STERGIOU E. Analytical Model for Performance Evaluation of Multilayer Multistage Interconnection Networks Servicing Unicast and Multicast Traffic by Partial Multicast Operation[J], Performance Evaluation, 2010, 67(10):959-976.

[4] MINKENBERG C, LUIJTEN R P, RODRIGUEZ G. On the Optimum Switch Radix in Fat Tree Networks[C]// Proceedings of the IEEE 12th

International Conference on High Performance Switching and Routing (HPSR), Cartagena, Spain: IEEE Press, 2011:44-51.
[5] NEWMAN P. Fast Packet Switching for Integrated Services (chapter four: Multistage Interconnection Networks)[D], Cambridge, UK: University of Cambridge, 1988.
[6] KOLIAS C and TOMKOS I. Switch Fabrics: Key Building Blocks for Switches and Routers[J], IEEE Circuits & Devices Magazine , 2005, 21(5):12-17.
[7] SMALL B A, BERGMAN K. Optimization of Multiple-Stage Optical Interconnection Networks, IEEE Photonics Technology Letters[J], 2006, 18(1):238-240.
[8] BRENNER M, ZIMMERMANN A. Analysis of Delay Time Distributions in Multistage Interconnection Networks Considering Multicast Traffic[C]// Proceedings of the seventh IEEE International Symposium on Network Computing and Applications, Cambridge, MA, USA: IEEE Press, 2008:236-239.
[9] VASILIADIS D C, RIZOS G E, VASSILAKIS C. Performance Study of Finite-Buffered Blocking Multistage Interconnection Networks Supporting natively 2-Class Priority Routing Traffic[J], Journal of Network and Computer Applications, 2013, 36(2):723-737.
[10] OHRING S R, IBEL M, Das S K, at al. On Generalized Fat Trees[C]// Proceedings of the 9th International Parallel Processing Symposium, Santa Barbara, CA, USA: IEEE Press, 1995:37-44.
[11] DAS S K, OHRING S R, IBEL M. Communication Aspects of Fat-Tree-Based Interconnection Networks for Multicomputers[C]// Proceedings of a DIMACS Workshop on Robust Communication Networks, Rhode Island, USA. 2000:35-60.
[12] NAVARIDAS J, MIGUEL-ALONSO J, FRANCISCO J R, at al. Reducing Complexity In Tree-Like Computer Interconnection Networks, Parallel computing[J], 2010, 36(2):71-85.
[13] LEISERSON E. Fat-Trees: Universal Networks for Hardware-Efficient Supercomputing, IEEE Transactions on Computers[J], 1985, 34(10): 892-901.
[14] PETRINI F, VANNESCHI M. K-ary N-trees: High Performance Networks for massively Parallel Architectures[C]// Proceedings of the 11th International Parallel Processing Symposium(IPPS'97), Geneva, Switzerland, 1997:87-93.
[15] ANJUM S, KHAN I A, ANWAR W, at al. A Scalable and Minimized Butterfly Fat Tree (SMBFT) Switching Network for on-Chip Communication[J], Research Journal of Applied Sciences, Engineering and Technology, 2012, 13(4):1997-2002.
[16] CHAO H J and LIU Bin, High Performance Switches and Routers, March 2007, Wiley-IEEE Press.
[17] PERATIKOU A, ADDA M. Optimising Extended Generalised Fat Tree Topologies[C]// Proceedings of the 17th International Conference on Distributed Computer and Communication Networks (DCCN2013), Moscow, Russia, 2013: 82-90.

Biographies

XU Zhanqi, received his B.S., M.S. and Ph.D. degrees in communication and electronic system from Xidian University, China in 1984, 1987, and 1997, respectively. Dr. XU had a one-year postdoctoral study at Hong Kong University of Science and Technology during the turn of this century. Since 2000, He has been with the state key laboratory on Integrated Services Networks (ISN), Xidian University, China, where he is currently a Professor. His interested areas include optical networks, space information networks, and communication networks modeling and performance evaluation.

WANG Chunting, received the Ph.D. degree in communication and information system from Xidian University, China in 2010. He currently serves as the deputy chief engineer in No.54 Institute of China Electronic Technology Corp. (CETC), China. His research areas include satellite communication and networking.

ZHOU Zhiqiang, received the M.S. degree in communication and electronic system from Wuhan Post & Telecommunication Research Institute, China in 1992. He is a senior engineer with the Fiberhome Technologies, Wuhan, China. His current interests include the implementation and performance evaluation of the large scale switching system.

HUANG Jiangjiang, received the B.S. and M.S. degrees in communication engineering from Xidian University, China in 2010 and 2013, respectively. His research interest includes broadband communication networks.

MA Tao, received the B.S. and M.S. degrees from Xian Jiaotong University, China, in 2005 and 2008 respectively, both in electrical engineering. He received the Ph.D. degree in Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, USA. He is currently a lecturer in state key laboratory on Integrated Services Networks (ISN), Xidian University, China. His research areas are Cross-layer design for QoS provisioning in wireless data networks, and multi-media distribution and 4G network evaluation.